


PEN

	
UNIVERSITAS PERPUSTAKAAN	
UNIVERSITAS JEMBER YOGYAKARTA	
PERPUSTAKAAN	
DATE	- 8 MAR 2011
INSTRUKSI	27/TIF/HO.3/Per/2011
REKAMASI	005.711/Ern/p
Subyek	: Programming



# LAPORAN PENELITIAN

## PEMBANGUNAN APLIKASI CLUSTERING DENGAN ALGORITMA IMMUNE K-MEANS



Oleh :

Dra. Ernawati, M.T.

Program Studi Teknik Informatika

Fakultas Teknologi Industri

Universitas Atma Jaya Yogyakarta

2011

## LEMBAR PENGESAHAN PENELITIAN

1. a) Judul penelitian : Pembangunan Aplikasi Clustering dengan Algoritma Immune K-means.  
b) Macam penelitian : Laboratorium
2. Personalia peneliti:
  - a) Nama : Dra. Ernawati, M.T.
  - b) Jenis kelamin : Perempuan
  - c) Usia saat pengajuan proposal : 41 tahun 5 bulan
  - d) Jabatan akademik, golongan : Lektor Kepala, IV/a
  - e) Bidang peminatan : Soft Computing
  - f) Fakultas, Program Studi : Teknologi Industri, Teknik Informatika
3. Jangka waktu penelitian : 6 bulan
4. Biaya yang disetujui : Rp. 3.025.000

Yogyakarta, 28 Februari 2011

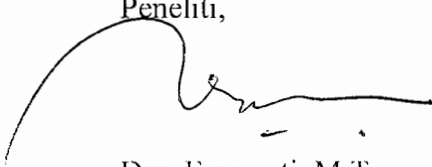
Mengetahui,

Kepala Laboratorium Komputasi FTI,



Patricia Ardanari, S.Si, M.T.

Peneliti,



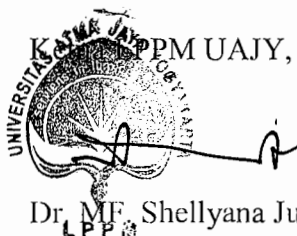
Dra. Ernawati, M.T.

Dekan Fakultas Teknologi Industri UAJY,



Ir. B. Kristyanto, M.Eng., Ph.D.

Kepala PPM UAJY,



Dr. MF. Shellyana Junaedi, S.E., M. Si.

## INTISARI

Clustering adalah suatu *task* dalam *data mining* yang mengelompokkan sejumlah data ke dalam klaster-klaster, sehingga setiap klaster akan berisi data yang semirip mungkin. Algoritma-algoritma clustering telah banyak dikemukakan, seperti algoritma K-means, Fuzzy K-means dan yang terbaru adalah algoritma Immune K-means, yang menggunakan prinsip sistem imun. *Artificial Immune System* adalah sistem komputasi yang diinspirasi oleh prinsip-prinsip dan proses-proses pada sistem kekebalan vertebrata, merupakan salah satu paradigma komputasional yang terbaru.

Pada penelitian ini dilakukan analisis dan perancangan aplikasi clustering untuk mengimplementasikan algoritma Immune K-means. Hasil clustering dengan metode Immune K-means akan dibandingkan dengan hasil clustering dengan algoritma K-means, menggunakan 40 sampel acak dari data lulusan.

Dari hasil pengamatan yang dilakukan, dapat disimpulkan bahwa karakteristik klaster yang terbentuk dari kedua algoritma tersebut memiliki karakteristik yang sama. Kelebihan dari algoritma K-means adalah komputasi relatif lebih sederhana, hasil cukup baik dilihat dari mean square error yang lebih kecil, namun memiliki kelemahan yaitu jumlah klaster ( $k$ ) harus ditentukan diawal, sedangkan algoritma Immune K-means dari sisi komputasi relatif lebih rumit, namun dengan adanya populasi B\_sel, dapat dipilih jumlah klaster yang terbaik, sehingga Immune K-means dapat berfungsi sebagai pembelajaran tak terbimbing maupun pembelajaran terbimbing (jika  $k$  ditentukan diawal).

Kata kunci : clustering, *Artificial Immune Systems*, Immune K-means

## KATA PENGANTAR

Puji syukur kepada Allah Yang Esa atas Berkat dan RahmatNya yang telah dianugerahkan sehingga penelitian dengan judul " Pembangunan Aplikasi Clustering dengan Algoritma Immune K-means" dapat selesai.

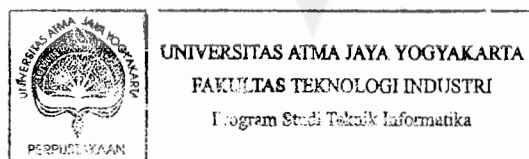
Saya haturkan terima kasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan penelitian ini, terutama kepada :

1. Bapak Ir. A. Koesmargono, M.Const.Mgt., Ph.D., selaku Rektor Universitas Atma Jaya Yogyakarta.
2. Ibu Dr. MF. Shellyana Junaedi, SE.,M.Si. , selaku Ketua LPPM UAJY.
3. Bapak Ir. B. Kristyanto, M. Eng., Ph.D., selaku Dekan Fakultas Teknologi Industri UAJY.
4. Ibu Patricia Ardanari S.Si., M.T., selaku Kepala Laboratorium Komputasi FTI.
5. Serta semua pihak yang tidak dapat kami sebutkan satu-persatu

Demikian laporan yang kami sampaikan, semoga laporan ini berguna.

Yogyakarta, Februari 2011

Peneliti



## DAFTAR ISI

Halaman Judul	
Halaman Pengesahan .....	i
Intisari .....	ii
Kata Pengantar .....	iii
Daftar Isi .....	iv
Daftar Tabel .....	v
Daftar Gambar .....	vi
<b>BAB 1. PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2. Perumusan Masalah .....	1
1.3. Tujuan Penelitian .....	2
1.4. Manfaat Penelitian .....	2
1.5. Batasan Masalah .....	2
<b>BAB 2. TINJAUAN PUSTAKA</b>	
2.1. Sistem Kekebalan ( <i>Immune System</i> ) .....	4
2.2. Artificial Immune System (AIS) .....	4
2.3. Clustering .....	7
2.4. Immune K-means .....	8
<b>BAB 3. METODOLOGI PENELITIAN</b>	
3.1. Jenis Penelitian .....	11
3.2. Metode Penelitian .....	11
3.3. Data Penelitian .....	11
<b>BAB 4. ANALISIS, PERANCANGAN DAN PEMBAHASAN</b>	
4.1. Analisis sistem .....	12
4.1.1. Lingkup Masalah .....	12
4.1.2. Perspektif Produk .....	12
4.1.3. Antarmuka Pemakai .....	12
4.1.4. Antarmuka Perangkat Keras .....	13
4.1.5. Antarmuka Perangkat Lunak .....	13
4.1.6. Fungsi Produk .....	13
4.1.7. Karakteristik Pengguna .....	13
4.1.8. Batasan-batasan .....	13
4.1.9. Kebutuhan Fungsional .....	14
4.1.10. Kebutuhan Data .....	15
4.2. Perancangan Sistem .....	15
4.2.1. Rancangan Arsitektur .....	15
4.2.2. Rancangan Proses .....	16
4.2.2. Rancangan Antarmuka .....	17
4.3. Pembahasan .....	19
<b>BAB 5. KESIMPULAN DAN SARAN</b> .....	25
<b>DAFTAR PUSTAKA</b> .....	26

## DAFTAR TABEL

Tabel 4.1. Data training .....	19
Tabel 4.2. Hasil Clustering menggunakan K_means dan Immune K-means ...	22



## DAFTAR GAMBAR

Gambar 4.1. Diagram konteks ClusIkMe .....	14
Gambar 4.2. DFD Level 1 ClusIkMe .....	14
Gambar 4.3. Arsitektur ClusIkMe .....	15
Gambar 4.4. Flowchart Proses Clustering dengan Immune K-means .....	16
Gambar 4.5. Halaman Utama ClusIkMe .....	17
Gambar 4.6. Halaman Proses Clustering dengan Immune K-Means .....	18
Gambar 4.7. Halaman Hasil Clustering ClusIkMe .....	18



## BAB I

### PENDAHULUAN

#### 1.1. Latar Belakang

Clustering adalah suatu *task* dalam *data mining* yang mengelompokkan sejumlah data ke dalam klaster-klaster, sehingga setiap klaster akan berisi data yang semirip mungkin. Algoritma-algoritma clustering telah banyak dikemukakan, seperti algoritma K-means, Fuzzy K-means dan lain-lain.

*Artificial Immune System* (AIS) adalah sistem komputasi yang diinspirasi oleh prinsip-prinsip dan proses-proses pada sistem kekebalan vertebrata. AIS muncul pertengahan tahun 1980-an dengan tulisan-tulisan dari Farmer, Packard, Perelson (1986) dan Barsini dengan Varela (1990), tetapi baru pertengahan tahun 1990-an AIS menjadi suatu subyek tersendiri. Selama dekade terakhir ini, AIS telah banyak menarik perhatian para peneliti. Telah banyak usaha yang dilakukan untuk mengembangkan algoritma-algoritma yang berguna untuk menyelesaikan masalah teknik, seperti pemrosesan dan analisis data. Salah satu algoritma clustering yang menggunakan prinsip sistem imun adalah algoritma Immune K-means.

Algoritma Immune K-means merupakan algoritma clustering yang mengkombinasikan prinsip Clonal Selection yang diinspirasi oleh teori imunologi dengan algoritma K-means, suatu algoritma clustering yang paling sederhana dan umum dikenal. Algoritma ini dikemukakan oleh Michal Bareta dan Tadeusz Burczynski yang dipublikasikan dalam jurnal tahun 2009. Karena algoritma ini merupakan suatu algoritma yang baru dan AIS juga merupakan suatu bidang kajian ilmu yang sedang banyak diperhatikan, maka pada penelitian ini akan dilakukan suatu studi untuk mengimplementasikan algoritma tersebut. Hasil clustering dengan metode Immune K-means akan dibandingkan dengan hasil clustering dengan algoritma K-means.

## 1.2. Perumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah :

1. Bagaimana membuat analisis dan rancangan aplikasi untuk mengimplementasikan algoritma clustering Immune K-means.
2. Bagaimana perbandingan hasil clustering antara algoritma K-means dan Immune K-means.

## 1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menganalisis dan merancang aplikasi clustering untuk mengimplementasikan algoritma clustering Immune K-means.
2. Membandingkan hasil clustering data menggunakan algoritma K-means dengan algoritma Immune K-means.

## 1.4. Manfaat Penelitian

Dari hasil penelitian ini didapat rancangan sebuah aplikasi clustering dengan algoritma Immune K-mean yang merupakan algoritma baru, pengembangan dari algoritma K-means dan hasil perbandingan algoritma Immune K-means dan algoritma K-means.

## 1.5. Batasan Masalah

Batasan masalah yang ditetapkan dalam penelitian ini adalah :

1. Algoritma clustering yang akan diimplementasikan adalah algoritma Immune K-means.

2. Data yang digunakan dalam pembahasan adalah data lulusan yang diambil dari [Ernawati, 2003]
3. Hasil clustering dari algoritma Immune K-means hanya akan dibandingkan dengan hasil clustering dengan algoritma K-means.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Sistem Kekebalan (*Immune System*)

Sistem kekebalan meliputi sel-sel, molekul-molekul dan organ-organ kompleks dengan peranan utama membatasi kerusakan pada *host organism* oleh patogen yang biasanya disebut antigen dengan memberikan respon kekebalan. Salah satu tipe respon adalah dengan proses mengeluarkan molekul-molekul antibodi oleh sel B atau limfosit B. Antibodi adalah molekul-molekul reseptor berbentuk Y yang terikat pada permukaan sel B dengan fungsi utama mengenali dan mengikat antigen melalui pencocokan yang saling melengkapi. Sel B melepaskan antibodi yang mengunci diri ke antigen spesifik, melumpuhkan patogen, dan menandainya untuk dihancurkan. Sel T (limfosit T) mengidentifikasi dan segera menghancurkan patogen.

Antibodi mengenali suatu bagian dari antigen yang disebut *epitope*-nya. Bagian antibodi yang bertanggung jawab mengenali suatu antigen disebut *paratope*, yang dikenal juga sebagai *V-region (variable region)*, berupa variabel karena dapat mengubah bentuknya untuk mencapai pencocokan yang lebih baik dengan antigen yang diberikan. Kekuatan dan kekhususan dari interaksi antibodi dan antigen diukur dengan dengan suatu ukuran yang disebut *affinity* dari kecocokannya.

#### 2.2. *Artificial Immune System (AIS)*

AIS didefinisikan oleh de Castro dan Timmis sebagai suatu sistem adaptif, diinspirasi oleh teori imunologi dan observasi fungsi kekebalan, prinsip dan model-model, yang diaplikasikan pada pemecahan masalah [Timmis dan Andrews]. Dalam ilmu komputer, AIS adalah sistem komputasi yang terinspirasi oleh prinsip-prinsip dan proses dari sistem kekebalan tubuh vertebrata. Sistem

kekebalan tubuh, khususnya vertebrata, adalah suatu sistem yang sangat kompleks dari interaksi sel-sel, organ dan mekanisme, yang tujuannya melindungi *host body* melawan bahaya, baik dari luar maupun dari dalam. Untuk melindungi *host body* dari bahaya-bahaya tersebut bukan tugas mudah, khususnya dalam lingkungan yang berubah. Sistem kekebalan perlu kemampuan untuk belajar mengenai ancaman baru, mengingat pengalaman sebelumnya dan mengembangkan respon yang khusus pada patogen yang berbeda. Algoritma AIS biasanya memanfaatkan karakteristik sistem kekebalan tubuh, dan yang paling menarik dari perspektif komputasi adalah *learning*, *diversity maintenance* dan *memory* untuk memecahkan masalah.

AIS muncul pertengahan tahun 1980-an dengan tulisan-tulisan dari Farmer, Packard, Perelson (1986) dan Barsini dengan Varela (1990), tetapi baru pertengahan tahun 1990-an AIS menjadi suatu subyek tersendiri. Selama dekade terakhir ini, AIS telah banyak menarik perhatian para peneliti. Telah banyak usaha yang dilakukan untuk mengembangkan algoritma-algoritma yang berguna untuk menyelesaikan masalah teknik, seperti pemrosesan dan analisis data. Forrest dan kawan-kawan serta Kephart dan kawan-kawan menerbitkan makalah pertama mereka mengenai AIS pada tahun 1994. Dasgupta melakukan studi ekstensif pada Algoritma Negative Selection. Hunt dan Cooke memulai pekerjaan pada model Immune Network pada tahun 1995, yang dilanjutkan dengan membuat beberapa perbaikan oleh Timmis dan Neal. De Castro dan Von Zuben serta Nicosia dan Cutello mengerjakan clonal selection, yang menjadi terkenal pada tahun 2002. Buku pertama mengenai Artificial Immune Systems diedit oleh Dasgupta pada tahun 1999.

Pengembangan utama dalam AIS berfokus pada tiga teori utama imunologi, yaitu clonal selection, immune network dan negative selection [Timmis dan Andrews]. Clonal selection mendeskripsikan fitur dasar dari respon kekebalan terhadap rangsangan antigenik. Algoritma ini mencontohkan ide bahwa hanya sel-sel yang mampu mengenali suatu rangsangan antigenik akan berkembangbiak, sehingga dipilih. Clonal selection beroperasi pada sel T dan sel B. Ketika suatu antibodi sangat cocok dengan suatu antigen, sel B teraktivasi dan

mulai berkembangbiak menghasilkan *clone-clone* sel B yang baru yang merupakan *copy* dari orangtuanya. *Clone-clone* tersebut kemudian mengalami *somatic hypermutation* (suatu mekanisme mutasi dengan laju tinggi) dan menghasilkan antibodi yang spesifik untuk melawan antigen. Algoritma clonal selection berfokus dari teori Darwin, dimana seleksi terinspirasi oleh kedekatan interaksi antigen-antibodi, reproduksi terinspirasi oleh pembelahan sel, dan variasi terinspirasi oleh somatic hypermutation. Algoritma clonal selection paling umum diterapkan untuk domain optimasi dan pengenalan pola. Teori Immune Network diusulkan oleh Niels Kaj Jerne pada tahun 1974, terinspirasi oleh teori jaringan idiotypic. Sistem kekebalan menjaga jaringan idiotypic dari interkoneksi sel-sel B untuk pengenalan antigen. Sel-sel ini menstimulasi dan menyembunyikan satu sama lain dengan cara tertentu untuk mencapai stabilitas jaringan. Dua sel B terhubung jika *affinity* keduanya melebihi suatu nilai ambang tertentu dan kekuatan hubungan berbanding langsung terhadap *affinity*-nya. Teori Immune Network telah digunakan dalam clustering, visualisasi data, kontrol, optimasi, dan berbagi properti dengan jaringan syaraf tiruan. Tujuan dari Negatif Selection adalah mentoleransi *self cells*. Hal ini berhubungan dengan kemampuan sistem kekebalan untuk mendeteksi antigen yang tidak dikenal, sementara itu tidak memberikan reaksi terhadap *self cells*. Algoritma ini terinspirasi oleh proses seleksi negatif dan positif yang terjadi selama pematangan sel T dalam timus, yang disebut toleransi sel T. Algoritma ini biasanya digunakan pada domain masalah klasifikasi dan pengenalan pola dimana ruang masalah dimodelkan dalam komplemen dari pengetahuan yang tersedia. Misalnya dalam kasus deteksi anomali, algoritma mempersiapkan satu set pola untuk dilatih pola normal (non-anomali) dan digunakan mendeteksi pola anomali.

Untuk mengimplementasikan AIS, ada empat keputusan yang harus dibuat, yaitu *Encoding*, *Similarity Measure*, *Selection* dan *Mutation*. Untuk mengimplementasikan AIS, perlu didefinisikan pengertian antigen dan antibodi dalam konteks domain aplikasi. Antigen dan antibodi merepresentasikan masalah yang akan diselesaikan dan suatu solusi. Antigen dan antibodi direpresentasikan atau dikodekan sebagai suatu string dari bilangan atau fitur, dengan panjang string menunjukkan jumlah variabel, posisi menunjukkan identifier dari variabel dan

nilai (bisa biner atau real) dari variabel. Setelah dilakukan *encoding*, maka ukuran kemiripan (*affinity*) harus dipilih. Ukuran kemiripan atau aturan kecocokan adalah salah satu hal yang paling penting dalam pemilihan disain dalam pengembangan algoritma AIS. Untuk *encoding* biner dapat menggunakan algoritma pencocokan dengan membandingkan posisi dari masing-masing bit, atau dapat juga menggunakan Hamming *distance*. Untuk *encoding* bilangan non biner seperti bilangan real dapat dihitung jarak diantara 2 string, misal menggunakan jarak Euclidean. Setelah *encoding* dan ukuran kemiripan ditentukan, maka algoritma akan melakukan seleksi dan mutasi berdasarkan ukuran kemiripan sampai kriteria berhenti terpenuhi.

### 2.3. Clustering

Teknik Clustering adalah suatu teknik dalam data mining yang mirip dengan klasifikasi. Perbedaan mendasar clustering dengan klasifikasi adalah kelompok-kelompok yang terbentuk (disebut klaster) belum didefinisikan terlebih dahulu. Klaster dibentuk dengan mencari kemiripan diantara data menurut karakteristik yang ditemukan dalam data aktual.

Ada banyak algoritma clustering, Arabie dan Hubert mengklasifikasikan algoritma-algoritma tersebut dalam 3 kategori berdasarkan tingkat kemungkinan overlap yaitu metode partisi, metode hirarki dan metode overlap penuh. Pendekatan yang paling banyak digunakan adalah metode hirarki dan metode partisi. Clustering dengan metode hirarki dilakukan dengan menggabungkan klaster-klaster kecil menjadi klaster yang lebih besar atau memecah klaster besar menjadi beberapa klaster yang lebih kecil. Hirarki dari klaster biasanya berupa sebuah pohon yang disebut dendogram. Hirarki dari klaster diciptakan melalui algoritma yang membangun algoritma klaster. Ada dua jenis utama clustering hirarki, yaitu *agglomerative* dan *divisive*. Teknik *agglomerative* mulai dengan  $n$  klaster. Klaster yang terdekat satu sama lain bergabung bersama untuk membentuk klaster yang lebih besar berikutnya. Penggabungan ini dilanjutkan sampai hirarki klaster hanya memuat sebuah klaster tunggal. Teknik *divisive*

menggunakan pendekatan yang berlawanan dari teknik *agglomerative*. Teknik ini dimulai dengan sebuah klaster besar, kemudian mencoba untuk membagi menjadi klaster-klaster yang lebih kecil sampai akhirnya setiap klaster hanya memuat sebuah data. Clustering dengan metode partisi secara langsung memecah kumpulan data kedalam klaster klaster yang saling asing. Suatu fungsi kriteria digunakan untuk meminimalisasi ketidakmiripan sampel-sampel dalam tiap klaster dan memaksimumkan ketidakmiripan sampel-sampel pada klaster yang berbeda. Salah satu metode partisi yang banyak dikenal adalah algoritma K-means.

Algoritma K-means adalah algoritma yang paling sederhana dan umum dikenal. Secara umum algoritma K-means dilakukan dengan langkah-langkah sebagai berikut :

1. Pilih jumlah klaster k.
2. Inisialisasi k pusat klaster.
3. Tempatkan setiap obyek ke klaster terdekat.
4. Hitung pusat klaster yang baru berdasarkan keanggotaan klaster yang baru.
5. Tempatkan kembali setiap obyek ke klaster terdekat, menggunakan titik pusat klaster yang baru. Jika titik pusat klaster tidak berubah lagi, maka proses klastering selesai, jika tidak, kembali ke langkah 3.

#### 2.4. Immune K-Means

Algoritma Immune K-means merupakan algoritma clustering yang mengkombinasikan prinsip Clonal Selection yang diinspirasi oleh teori imunologi dengan algoritma K-means. Algoritma ini dikemukakan oleh Michal Bareta dan Tadeusz Burczynski yang dipublikasikan dalam jurnal tahun 2009. Secara umum algoritma Immune K-means dilakukan dengan langkah-langkah sebagai berikut :



1. Bentuk suatu populasi awal dari sel B sebagai suatu vektor bernilai real secara random sebagai suatu subset dari *training set*.
2. Untuk setiap antigen  $a_i$  (suatu sampel dalam *training set*), temukan sel B terdekatnya.
3. Untuk setiap sel B, bentuk *clone*-nya sebagai suatu vektor rata-rata dari semua antigen di mana suatu sel B yang diberikan adalah tetangga terdekat. Jika suatu *clone* sama dengan *parent*, mutasi *clone* dengan menambahkan suatu nilai random dengan range  $[-mut, mut]$  ke setiap dimensi. Ini menghindarkan *clone* dari tempat tanpa klaster. Tambahkan *clone* ke populasi dari sel B.
4. Untuk setiap antigen, cari sel B baru terdekat.
5. Hitung tingkat stimulasi dari setiap sel B. Tingkat stimulasi untuk sel B ke-j dihitung menggunakan persamaan :

$$\text{Tingkat stimulasi}(B_j) = \frac{\sum_{a_i \in NN(B_j)} e^{-\beta(euc\_disc(B_j, a_i))}}{\sum_{a_i \in NN(B_j)} 1}$$

dengan :

- $NN(B_j)$  adalah suatu subset dari antigen di mana  $B_j$  adalah tetangga terdekat.
- $\beta$  adalah konstanta positif
- $euc\_disc$  adalah jarak euclidean

6. Urutkan sel B dalam urutan menurun menurut tingkat stimulasi
7. Lakukan supresi
8. Ulangi langkah 3-7 sampai kondisi berhenti terpenuhi.

Keuntungan terbesar dari algoritma ini adalah bahwa setiap sel B menciptakan hanya satu *clone* dalam setiap iterasi. Konsep supresi berdasarkan pada *usefulness* dari sel yang diberikan. Supresi dilakukan dengan cara sebagai

berikut : mulai dari  $B_j$  yang paling kecil terstimulasi, untuk setiap  $a_i \in \text{NN}(B_j)$ , cari  $B_{\text{new}}$  sebagai tetangga terdekat baru diantara sel  $B$  (kecuali  $B_j$ ) dari  $a_i$  dan hitung jaraknya ke  $B_{\text{new}}$ . Jika kondisi

$$(\text{jarak baru} - \text{jarak lama}) \leq \alpha$$

terpenuhi untuk setiap  $a_i \in \text{NN}(B_j)$ , buang  $B_j$  dan ingat tetangga terdekat baru untuk setiap  $a_i \in \text{NN}(B_j)$ . Nilai jarak baru dan jarak lama adalah jarak Euclidean antara antigen  $a_i$  yang diberikan dengan  $B_j$  dan jarak Euclidean antara antigen  $a_i$  yang diberikan dengan  $B_{\text{new}}$ .

Sampel *training* dinormalisasi ke range  $[0,1]$ . Parameter beta tidak secara signifikan mempengaruhi evolusi, sehingga diset =1. Parameter mut menurun ke 0 selama iterasi berlangsung, dengan nilai awal yang kecil (misal 0,05). Parameter yang paling penting adalah alpha, yang menunjukkan perubahan maksimum yang diijinkan pada jarak ke tetangga terdekat untuk setiap antigen sementara mencoba membuang sel  $B$  yang diberikan. Nilai awal parameter ini kecil (misal 0,001), kemudian bertambah besar selama iterasi. Nilai awal yang kecil memungkinkan menemukan struktur data pada permulaan evolusi. Nilai akhir tergantung pada training set dan dapat dianggap sebagai ukuran densitas dari populasi sel-sel  $B$ . Semakin besar nilai akhir alpha, semakin sedikit sel-sel  $B$  yang akan ada dalam populasi akhir.

## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Jenis Penelitian**

Penelitian yang dilakukan adalah penelitian terapan yang berkaitan dengan salah satu paradigma dalam kecerdasan komputasional, yaitu Artificial Immune System. Penelitian ini dilakukan dengan menganalisis dan merancang aplikasi clustering untuk mengimplementasikan algoritma clustering Immune K-means, yang merupakan algoritma baru hasil pengembangan dari algoritma K-means. Selain itu akan dibandingkan hasil clustering menggunakan algoritma Immune K-means dengan algoritma K-means.

#### **3.2. Metode Penelitian**

Metode yang digunakan dalam penelitian ini adalah sebagai berikut :

##### **a. Studi pustaka**

Studi pustaka dilakukan untuk mengkaji dan memahami mengenai Artificial Immune System serta algoritma Immune K-means.

##### **b. Analisis dan Perancangan Perangkat lunak**

Analisis dan perancangan perangkat lunak dilakukan sebagai dasar pembangunan perangkat lunak.

##### **c. Analisis data**

Analisis data dilakukan untuk mendapatkan hasil perbandingan antara algoritma Immune K-means dan algoritma K-means.

#### **3.3. Data Penelitian**

Untuk analisis hasil, data yang digunakan adalah data sekunder dari jurnal hasil penelitian yang pernah dilakukan peneliti sebelumnya [Ernawati, 2003].

## **BAB IV**

### **ANALISIS, PERANCANGAN dan PEMBAHASAN**

#### **4.1. Analisis Sistem**

Pada penelitian ini akan dianalisis dan dirancang perangkat lunak yang digunakan untuk melakukan clustering menggunakan algoritma Immune K-means. Sebagai studi kasus data yang akan dianalisis adalah data lulusan dari [Ernawati, 2003]. Perangkat lunak ini diberi nama ClusIkMe.

##### **4.1.1. Lingkup masalah**

Perangkat lunak ClusIkMe dibangun dengan tujuan untuk melakukan clustering data lulusan dengan metode Immune K-means sehingga dapat dibandingkan hasil clustering antara metode Immune K-means dengan metode K-means.

##### **4.1.2. Perspektif Produk**

Perangkat lunak ClusIkMe merupakan perangkat lunak yang membaca data yang akan dicluster, dalam hal ini adalah data lulusan. Berdasarkan data tersebut, dilakukan clustering dengan metode Immune K-means, sehingga diketahui bagaimana klaster yang terbentuk.

##### **4.1.3. Antarmuka Pemakai**

Interaksi sistem dengan pengguna dilakukan pada monitor standar dengan spesifikasi 800x600 pixel. Menu yang ditampilkan adalah menu window, berupa hasil clustering yang dilakukan dengan metode Immune K-means.

#### 4.1.4. Antarmuka Perangkat Keras

Piranti antarmuka perangkat keras yang digunakan perangkat lunak ClusIkMe adalah mouse dan keyboard.

#### 4.1.5. Antarmuka Perangkat Lunak

Perangkat lunak yang dibutuhkan untuk menjalankan perangkat lunak ClusIkMe adalah Microsoft Windows sebagai sistem operasi.

#### 4.1.6. Fungsi Produk

Fungsi produk yang dimiliki perangkat lunak ClusIkMe adalah :

1. Fungsi Clustering untuk melakukan proses clustering dan menampilkan hasil clustering dengan algoritma Immune K-means.
2. Fungsi Exit, untuk keluar dari aplikasi.

#### 4.1.7. Karakteristik Pengguna

Pengguna perangkat lunak ClusIkMe yaitu siapa saja yang ingin melakukan analisis clustering dengan algoritma Immune K-means dengan syarat dapat mengoperasikan komputer dan mengerti konsep clustering.

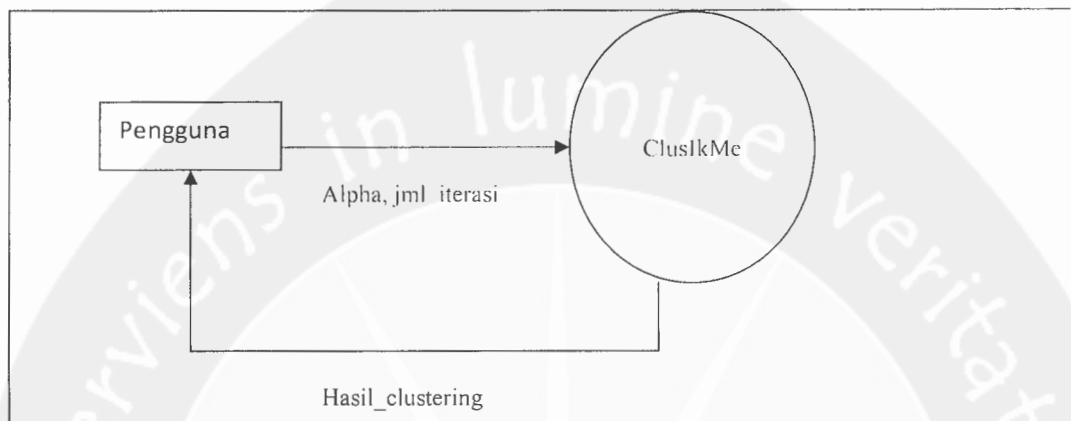
#### 4.1.8. Batasan-batasan

Adapun batasan dalam pembangunan perangkat lunak ini adalah data yang akan dicluster tersedia dalam bentuk file Access.

#### 4.1.9. Kebutuhan Fungsional

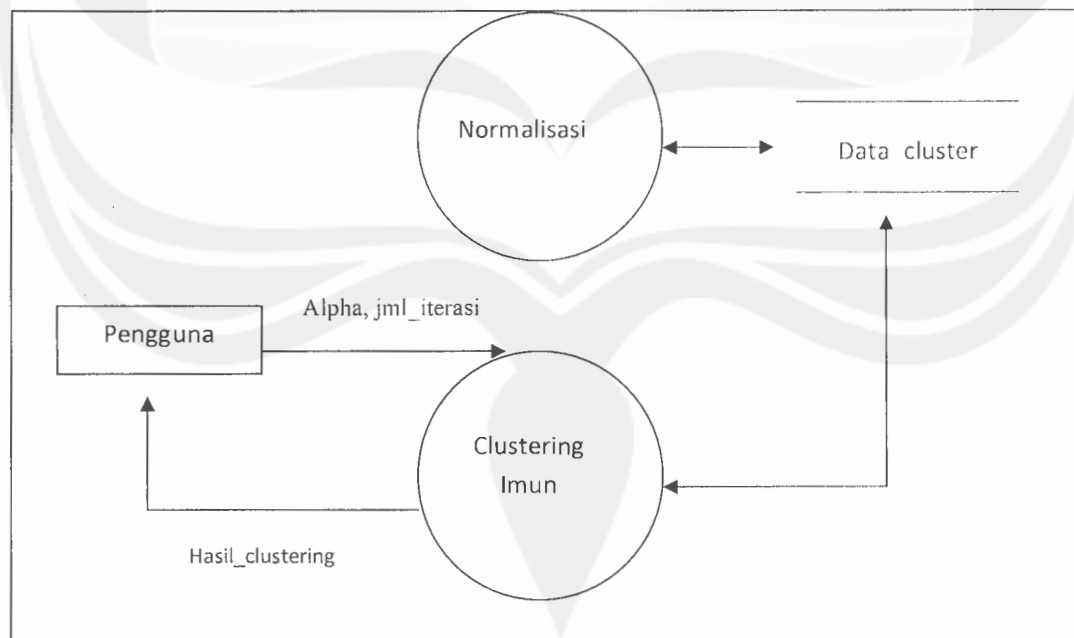
Kebutuhan fungsionalitas perangkat lunak ClusIkMe dianalisis menggunakan Data Flow Diagram (DFD) berikut :

##### 4.1.9.1. Diagram konteks



Gambar 4.1. Diagram konteks ClusIkMe

##### 4.1.9.2. DFD Level 1



Gambar 4.2. DFD Level 1 ClusIkMe

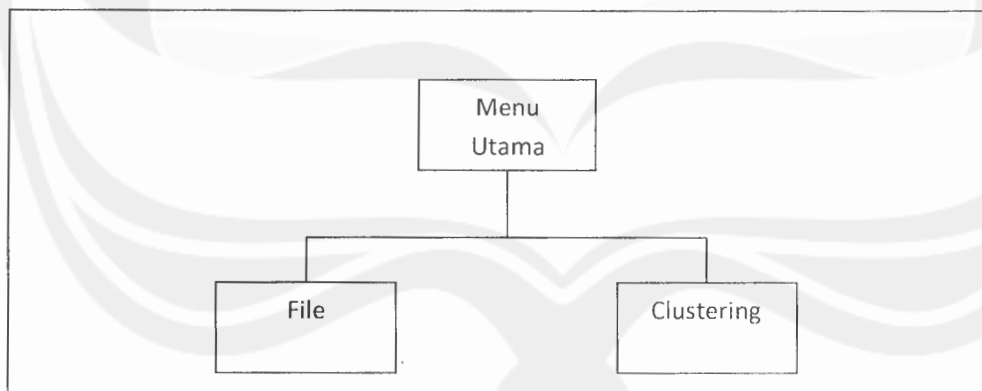
#### 4.1.10. Kebutuhan Data

Data yang dibutuhkan dan hasil pengolahan ClusIkMe disimpan dalam tabel datacluster (no, IPK, LamaStudi, Norm\_IPK, Norm\_LamaStudi, JarakMin, JarakBaru, Cluster). No sebagai primary key, IPK memuat data IPK lulusan, LamaStudi berisi data lama studi lulusan, Norm\_IPK dan Norm\_LamaStudi masing-masing memuat data hasil normalisasi dari data IPK dan LamaStudi, JarakMin menyimpan data jarak ke Sel B terdekat, JarakBaru digunakan untuk menyimpan jarak terdekat dari suatu Sel B ke Sel B baru lainnya jika sel B terdekat mengalami supresi, dan Cluster memuat data nama hasil clustering.

#### 4.2. Perancangan Sistem

Pada subbab ini akan ditampilkan rancangan arsitektur, rancangan antarmuka dan rancangan proses perangkat lunak ClusIkMe.

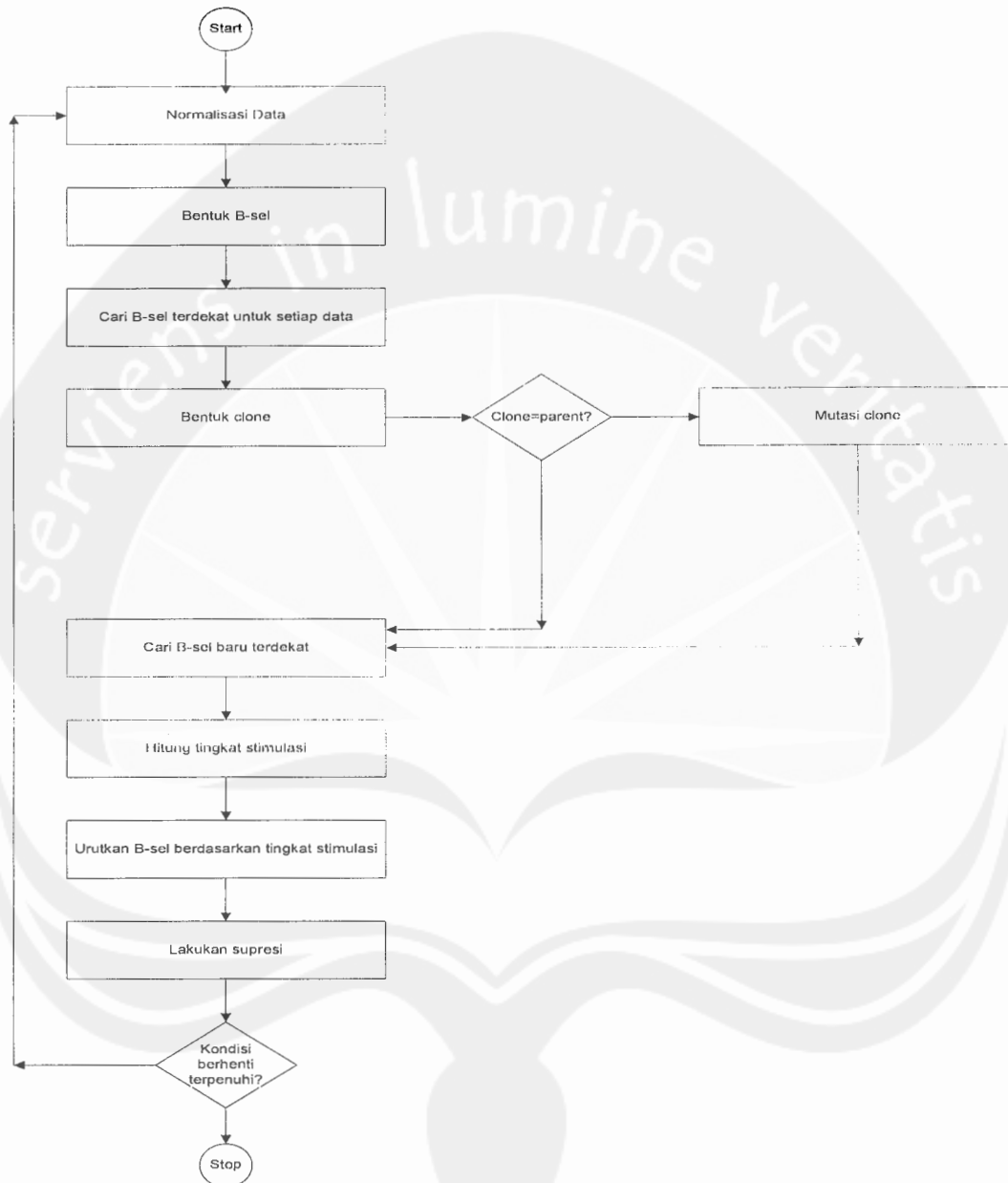
##### 4.2.1. Rancangan Arsitektur



Gambar 4.3. Arsitektur ClusIkMe

#### 4.2.2. Rancangan Proses

Berikut ini adalah flowchart yang menggambarkan rancangan proses yang akan dilakukan untuk proses clustering dengan algoritma Immune K-means,

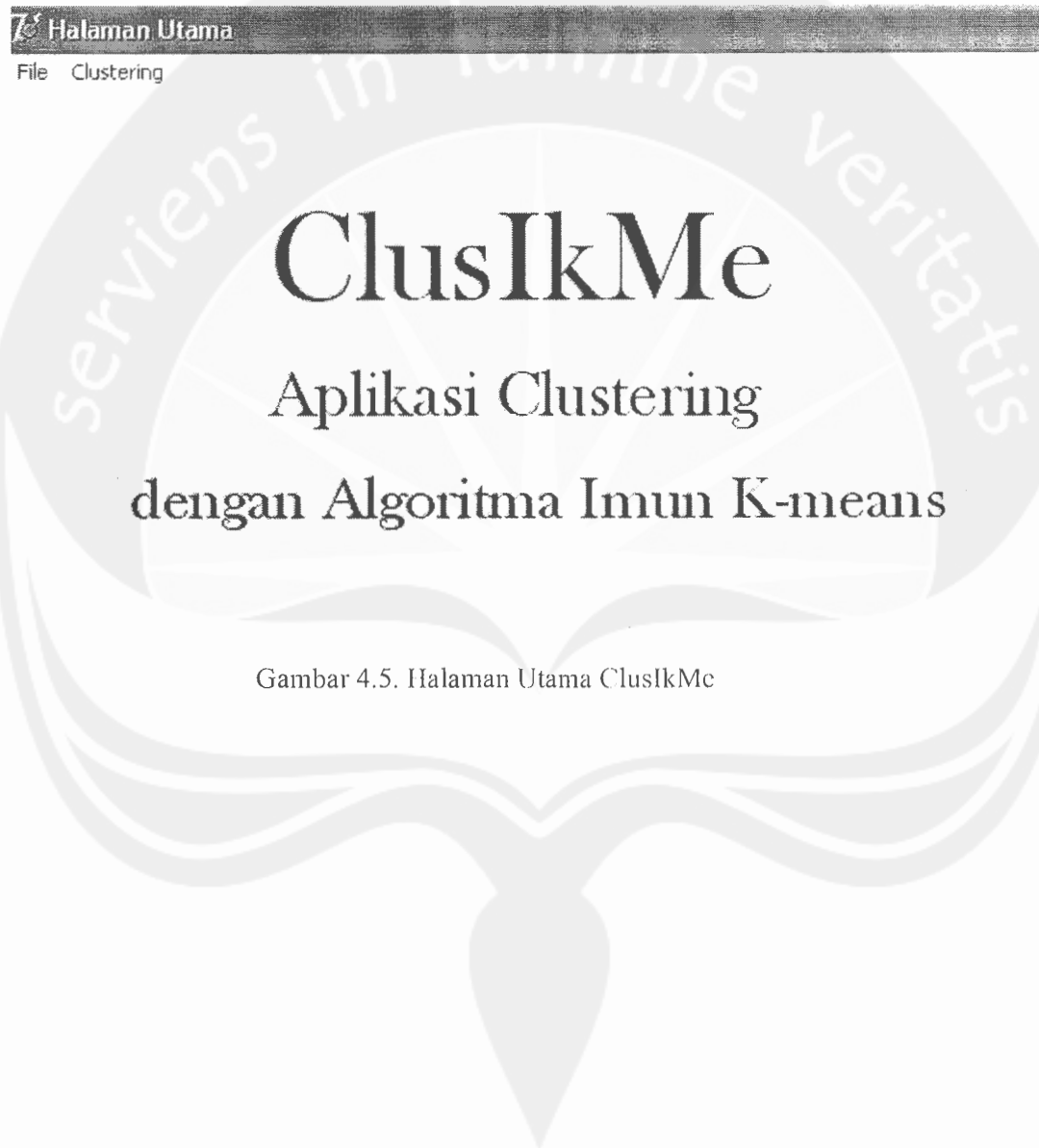


Gambar 4.4. Flowchart Proses Clustering dengan Immune K-means



#### 4.2.3. Rancangan Antarmuka

Berikut ini akan disajikan gambar rancangan form sebagai antarmuka ClusIkMe. Gambar 4.5. merupakan form halaman utama aplikasi, gambar 4.6. adalah form halaman antarmuka untuk menerima inputan parameter dan gambar 4.7. adalah form untuk menampilkan hasil clustering.



Gambar 4.5. Halaman Utama ClusIkMe

Masukkan nilai :

Alpha = 0.01

Jumlah iterasi = 10

OK

Gambar 4.6. Halaman Proses Clustering dengan Immunc K-Means

Masukkan nilai :

Alpha = 0.01

Jumlah iterasi = 10

OK

no	IPK	Lamastudi	cluster
1	2.75	5.5	1
2	2.75999999046326	5.5	1
3	2.599999990463257	5.82999992370605	1
4	2.82999992370605	5.82999992370605	1
5	2.80999994277954	5.82999992370605	1

Gambar 4.7. Halaman Hasil Clustering ClusIkMe

#### 4.3. Pembahasan

Untuk pembahasan ini, digunakan data lulusan yang diambil dari penelitian [Ernawati, 2003]. Data tersebut berupa 40 sampel acak yang memuat IPK dan lama studi lulusan., yang dijadikan sebagai *sampel training*. Sebelum digunakan untuk *training*, data tersebut perlu dinormalisasi ke dalam range  $[0,1]$ . Data sebelum dan sesudah normalisasi disajikan dalam Tabel 4.1.

Tabel 4.1. Data training

No Sampel	IPK (Xi)	LamaStudi (Yi)	Norm_IPK	Norm_LamaStudi
1	2.75	5.50	0.29	0.64
2	2.76	5.50	0.29	0.64
3	2.60	5.83	0.20	0.58
4	2.83	5.83	0.33	0.58
5	2.81	5.83	0.32	0.58
6	2.49	6.17	0.14	0.52
7	2.55	6.17	0.17	0.52
8	2.53	6.17	0.16	0.52
9	2.60	6.83	0.20	0.41
10	2.45	7.25	0.11	0.34
11	3.49	4.85	0.69	0.76
12	3.42	4.83	0.67	0.76
13	3.11	5.17	0.49	0.70
14	2.76	5.17	0.29	0.70

Tabel 4.1. Data training (lanjutan)

15	2.63	5.83	0.22	0.58
16	2.62	6.42	0.21	0.48
17	2.63	6.42	0.22	0.48
18	2.62	6.67	0.21	0.44
19	3.37	4.17	0.64	0.87
20	2.96	4.50	0.41	0.82
21	2.79	4.50	0.31	0.82
22	3.21	4.83	0.55	0.76
23	3.05	4.83	0.46	0.76
24	2.77	6.25	0.30	0.51
25	3.13	5.42	0.50	0.66
26	3.72	4.25	0.84	0.86
27	3.10	5.17	0.49	0.70
28	2.92	5.42	0.38	0.66
29	3.12	5.42	0.50	0.66
30	2.94	5.42	0.39	0.66
31	2.98	5.42	0.42	0.66
32	2.86	3.67	0.35	0.96
33	2.93	7.83	0.39	0.23
34	2.67	6.83	0.24	0.41

Tabel 4.1. Data training (lanjutan)

35	2.84	5.83	0.34	0.58
36	2.77	4.83	0.30	0.76
37	2.37	8.17	0.07	0.17
38	2.72	4.17	0.27	0.87
39	2.84	5.25	0.34	0.69
40	2.68	5.33	0.25	0.67

Agar dapat memilih solusi yang terbaik, perlu ditentukan *similarity measure*, yaitu ukuran kemiripan atau aturan kecocokan antara data sampel sebagai antigen dengan sel B yang dihasilkan sebagai antibodi. Ukuran kemiripan yang dipakai dalam penelitian ini adalah jarak Euclidean. Kriteria seleksi yang diterapkan adalah jarak terkecil. Sedangkan mutasi dilakukan dengan membentuk *clone* untuk sel B dengan mengambil vektor rata-rata dari semua antigen di mana sel B tersebut sebagai tetangga terdekat.

Untuk pembahasan ini, dilakukan *training* terhadap data menggunakan nilai awal  $\alpha=0,01$ , dengan kriteria berhenti sampai didapat 3 klaster (berdasar penelitian sebelumnya,  $k$  terbaik = 3). Hasil clustering dengan Immune K-means dan K-means dengan  $k=3$  dapat dilihat pada table 4.2.

Tabel 4.2. Hasil Clustering menggunakan K\_means dan Immune K-means

No Sampel	IPK (Xi)	LamaStudi (Yi)	Cluster (K-means)	Cluster (Immune K-means)
1	2.75	5.50	1	1
2	2.76	5.50	1	1
3	2.60	5.83	3	1
4	2.83	5.83	1	1
5	2.81	5.83	1	1
6	2.49	6.17	3	1
7	2.55	6.17	3	1
8	2.53	6.17	3	1
9	2.60	6.83	3	1
10	2.45	7.25	3	1
11	3.49	4.85	2	2
12	3.42	4.83	2	2
13	3.11	5.17	2	2
14	2.76	5.17	1	1
15	2.63	5.83	3	1
16	2.62	6.42	3	1
17	2.63	6.42	3	1
18	2.62	6.67	3	1

Tabel 4.2. Hasil Clustering menggunakan K\_means dan Immune K-means  
(lanjutan)

19	3.37	4.17	2	3
20	2.96	4.50	1	3
21	2.79	4.50	1	1
22	3.21	4.83	2	3
23	3.05	4.83	2	3
24	2.77	6.25	3	1
25	3.13	5.42	2	3
26	3.72	4.25	2	2
27	3.10	5.17	2	3
28	2.92	5.42	1	1
29	3.12	5.42	2	3
30	2.94	5.42	1	1
31	2.98	5.42	1	1
32	2.86	3.67	1	3
33	2.93	7.83	3	1
34	2.67	6.83	3	1
35	2.84	5.83	1	1
36	2.77	4.83	1	1
37	2.37	8.17	3	1
38	2.72	4.17	1	1

Tabel 4.2. Hasil Clustering menggunakan K\_means dan Immune K-means  
(lanjutan)

39	2.84	5.25	1	1
40	2.68	5.33	1	1

Titik pusat klaster yang diperoleh dengan algoritma K\_means adalah : klaster 1 : (2,83 ; 5,14), klaster 2 ( 3,27 ; 4,89) dan klaster 3 (2,60 ; 6,63), dengan mean square error 16,92. Titik pusat klaster yang diperoleh dengan algoritma Immune K\_means adalah : klaster 1 : (2,71 ; 5,96), klaster 2 ( 3,54 ; 4,64) dan klaster 3 (3,10 ; 4,80), dengan mean square error 62,30. Dari hasil yang diperoleh tersebut, dapat disimpulkan bahwa ketiga klaster yang terbentuk dari kedua algoritma tersebut memiliki karakteristik yang sama. Dari algoritma K-means didapat klaster dengan karakteristik sebagai berikut : klaster 1 memuat lulusan dengan IPK sedang dan masa studi sedang, klaster 2 memuat lulusan dengan IPK tinggi dan masa studi cepat dan klaster 3 memuat lulusan dengan IPK rendah dan masa studi lama. Dari algoritma Immune K-means didapat klaster dengan karakteristik sebagai berikut : klaster 1 memuat lulusan dengan IPK rendah dan masa studi lama, klaster 2 memuat lulusan dengan IPK sangat tinggi dan masa studi cepat dan klaster 3 memuat lulusan dengan IPK sedang dan masa studi sedang.

Kelebihan dari algoritma K-means adalah komputasi relatif lebih sederhana, hasil cukup baik dilihat dari mean square error yang lebih kecil, namun memiliki kelemahan yaitu jumlah klaster (k) harus ditentukan diawal, sedangkan algoritma Immune K-means dari sisi komputasi relatif lebih rumit, namun dengan adanya populasi B\_sel, dapat dipilih jumlah klaster yang terbaik, sehingga Immune K-means dapat berfungsi sebagai pembelajaran tak terbimbing maupun pembelajaran terbimbing (jika k ditentukan diawal).



## **BAB VI**

### **KESIMPULAN DAN SARAN**

Berdasarkan analisis dan perancangan sistem yang telah dilakukan serta implementasi algoritma Immune K-means pada data lulusan yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut :

1. Klaster yang terbentuk dari algoritma K-means dan Immune k-means memiliki karakteristik yang sama.
2. Komputasi pada Algoritma K-means relatif lebih sederhana dibandingkan Algoritma Immune K-means.
3. Algoritma Immune K\_means tidak mensyaratkan harus menentukan nilai k (banyaknya klaster di awal), sehingga dapat berfungsi untuk pembelajaran terbimbing maupun pembelajaran tak terbimbing.

Sebagai saran untuk pengembangan dari penelitian ini adalah mengimplementasikan pada data dengan jumlah lebih besar, sehingga dapat diamati keuntungan dari proses mutasi dan seleksi menggunakan metode supresi.

## DAFTAR PUSTAKA

1. Bareta, Michal & Burczynski, Tadeusz, 2009, Immune K-means and Negative Selection Algorithms for Data Analysis, Information Sciences 179 (2009).
2. De Castro, L. N. & Von Zuben F. J., 2000, An Evolutionary Immune Network for Data Clustering, Proc. of the IEEE SBRN (Brazilian symposium on artificial Neural Network).
3. Dunham, Margaret H., 2003, Data Mining Introductory and Advanced Topics, Pearson Education, Inc.
4. Ernawati, 2003, Aplikasi Algoritma Clustering untuk Klasifikasi Lulusan, Jurnal Teknologi Industri, Vol VII, no 2 , halaman 89-98.
5. Engelbrecht, 2007, Computational Intelligence An Introduction, Second Edition, John Wiley & Sons, Ltd,
6. Timmis, Jon & Andrews, Paul, A Beginners Guide to Artificial Immune Systems

